Automated Classification of Beaked Whales and Other Small Odontocetes in the Tongue of the Ocean, Bahamas

Susan Jarvis, Nancy DiMarzio, Ronald Morrissey and David Morretti
Naval Undersea Warfare Center Division, Newport, RI 02841

***Abstract*: Navy sonar has recently been associated with a number of marine mammal stranding events[1]. Beaked whales have been the predominant species involved in a number of these strandings. Monitoring and mitigating the effects of anthropogenic noise on marine mammals are active areas of research. Key to both monitoring and mitigation is the ability to automatically detect and classify the animals, especially beaked whales. This paper presents a novel support vector machine based methodology for automated species level classification of small odontocetes. To date, the algorithm presented has been trained to differentiate the click vocalizations of Blainville's beaked whales (*Mesoplodon densirostris*) from the clicks produced by delphinids and from man-made sounds. The automated classification capability compliments the detection and tracking tools already developed through ONR funding for the monitoring and localization of whales at the Atlantic Undersea Test and Evaluation Center, Andros Island, Bahamas.**

I. INTRODUCTION

Until very recently, little was know about beaked whale vocalizations. However, starting with the definitive recording of beaked whale clicks by Tyack, Johnson, et al. (using non-invasive DTAG's) [1, 2] and continuing with the visually verified recording of beaked whales and other small odontocete vocalizations at AUTEC [3] there is now sufficient labeled data to develop automated classification algorithms. This paper investigates the application of a novel class-specific support vector machine to the classification of vocalizations from beaked whales and small odontocetes.

At a basic level, a classification system is one that assigns the current input $\boldsymbol{x}$ membership in to one of k known classes according to some set of decision metrics or functions. In general, $\boldsymbol{x}$ is a multivariate random variable such that $\boldsymbol{x} \sim P(\boldsymbol{x})$. For example, popular maximum likelihood classifiers [4], assign an input data vector $\boldsymbol{x}$ membership in one of k possible class hypotheses $\{H_1, \ldots H_j \ldots H_k\}$ according to the probabilistic rule $j^* = arg\ max(p(H_j|\mathbf{x}))$. This is equivalently written as $j^* = arg\ max(p(\mathbf{x}|H_j)p(H_j))$ after applying Bayes rule. Theoretically, a maximum likelihood (ML) classifier is optimal in that it offers the lowest probability of error of any classifier [4]. However, in practice, it can be difficult to attain this optimal performance because the multidimensional probability density functions $p(\mathbf{x}|H_j)$ are unknown and must be estimated from training data. The amount of training data required to accurately estimate $p(\mathbf{x}|H_j)$ grows exponentially with the dimension of $\boldsymbol{x}$. This is problematic because the collection of labeled training data is usually difficult, time consuming and expensive.

Statistical learning theory [5,6] represents a different paradigm for learning than the classical ML methods presented above. Statistical learning theory advocates solving specific problems directly vice solving more general problems as an intermediate step [5]. That is, if there are limited data available to train a classifier then the best course of action is to estimate a decision boundary directly from the data. This is in contrast to classical ML inference where the data are used to estimate parameters of density functions and then the PDFs are used to form decision boundaries.

II. DISCUSSION

One of the corner stones of statistical learning theory is the principle of structured risk minimization (SRM). Using the SRM principle, Vapnik developed a bound on the risk of classification error for a given decision function f given the empirical risk (training error) $R_{emp}(f)$ associated with the function, the training set size m, and the capacity h of the hypothesis space in which the decision function resides [6]. This bound (1) is often referred to as the guaranteed risk, and is independent of the underlying distribution of the data. According to the SRM principle, the smallest bound on classification error is achieved by minimizing training error while using the function hypothesis space of the smallest capacity [5,6].

$$R[f] \le R_{emp}[f] + \sqrt{\frac{1}{m}\left(h\left(\ln\frac{(2m)}{h}+1\right)+\ln\frac{4}{\delta}\right)} \quad [6] \qquad (1)$$

Support vector methods (or support vector machines, SVM) are a rich family of learning algorithms based on statistical learning theory. SVM's were originally developed to solve binary classification problems of the following type: Given a set of empirical data $\{(\mathbf{x}_1, y_1) \ldots (\mathbf{x}_i, y_i) \ldots (\mathbf{x}_m, y_m)\}$ where each (multidimensional) input example $\mathbf{x}_i$ drawn from X is associated with classification label $y_i = \pm 1$, determine the decision function that maps any new $\boldsymbol{x}$ drawn from X to $y = \pm 1$ that minimizes risk of misclassification [5]. In short, SVMs implement the SRM principle.

[1] *Letter from the National Resource Defense Council, Oceans Futures Society, International Fund for Animal Welfare and the Humane Society of the United States to Hon. Gordon R. England, Secretary of the Navy, 14 July 2002.*

1-4244-0115-1/06/$20.00 ©2006 IEEE

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01 SEP 2006	**N/A**	**-**

4. TITLE AND SUBTITLE	
Automated Classification of Beaked Whales and Other Small Odontocetes in the Tongue of the Ocean, Bahamas	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Undersea Warfare Center Division, Newport, RI 02841**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S)
	11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM002006. Proceedings of the MTS/IEEE OCEANS 2006 Boston Conference and Exhibition Held in Boston, Massachusetts on September 15-21, 2006, The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**6**	

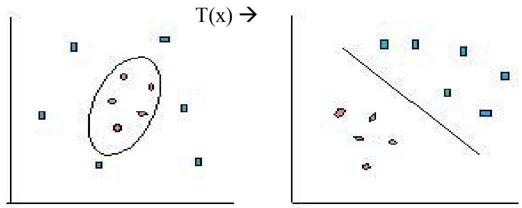


Figure 1: A notional view of a SVM [6]. a) Training data drawn from $\boldsymbol{x}$ shows two classes. b) Transformation T($\boldsymbol{x}$) maps the training data to a higher dimensional space where the optimal separating hyperplane is found. The hyperplane in the higher dimensional space corresponds to a nonlinear decision boundary in the input space.

SVM's use the existence of a unique optimal hyperplane which separates the two classes in some feature space (fig. 1). The SVM that implements the optimal hyperplane while maximizing the separation (margin) between the two classes will have the lowest risk of test error [5]. This optimal separating hyperplane is realized as

$$f(\boldsymbol{x}) = \sum_{i=1}^{m} \alpha_i y_i G(\boldsymbol{x}, \boldsymbol{x}_i) + b \qquad (2)$$

where G is a kernel mapping and b is an offset. The weights α_i for a "soft" margin SVM classifier [6] are found through

$$\max W(\boldsymbol{\alpha}) = \sum_{i=1}^{m} \alpha_i - \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j G(\boldsymbol{x}, \boldsymbol{x}_i) \qquad (3)$$

subject to $0 \le \alpha_i \le C/m,\ i = 1,2,...,m$ and $\sum_{i=1}^{m} \alpha_i y_i = 0.$

The constant C controls the degree of "slack" in the threshold optimization. Large C corresponds to more rigid separation of the classes and less tolerance for class overlap in the training data. Smaller C allows for more class overlap in the training data [7]. The optimization problem (3) is commonly solved using quadradic programming techniques [6, 8].

While SVM's were originally formulated for binary classification, many real world problems involve more than two classes. As a result, a number of methods have been developed for applying SVM's to multi-class problems. These methods tend to follow one of three basic approaches. The first approach is to form k binary "one-against-the-rest" classifiers (where k is the number of class labels) and choose the class whose decision function is maximized [5]. The second approach is to form all $k(k-1)/2$ pairwise binary classifiers and choose the class whose pairwise decision functions are maximized [9]. The third approach is to reformulate the objective function of the SVM for the multi-class case such that the decision boundaries for all classes are optimized jointly [10, 11].

In this paper we present a new type of multi-class support vector classifier called the class-specific SVM (CS-SVM). The new classifier consists of k binary SVM's where each SVM discriminates between one of k classes of interest and a common reference class. The class whose decision function is maximized with respects to the reference class is selected. The CS-SVM extends the concept of exploiting class-specific features as proposed by other researchers for maximum likelihood classifiers [4,12] and neural networks [13] to the multi-class SVM problem.

Many applications involve the classification of signals which are set in additive noise. In that case, the problem is not to differentiate between two or more of k signals but to differentiate between one of k signals and noise. The input vectors for such problems are actually of the form $\boldsymbol{x}_u = \boldsymbol{s}_u + \boldsymbol{n}$, for $u = 1,2, ... k$. Currently, SVM's are designed assuming the classification problem is distinguishing $\boldsymbol{x}_u = \boldsymbol{s}_u$ from $\boldsymbol{x}_v = \boldsymbol{s}_v$ Any noise in $\boldsymbol{x}$ is assumed to be accommodated by allowing "slack variables" in the hyperplane optimization [6].

The CS-SVM expressly acknowledges the presence of the noise by treating it as a reference class. For a single class, the classification problem reduces to a decision as to whether signal $\mathbf{s}$ is present or not. That is, $y = sgn(f(\boldsymbol{x})) = +1$ when $\boldsymbol{x} = \boldsymbol{s} + \boldsymbol{n}$ and $y = sgn(f(\boldsymbol{x})) = -1$ when $\boldsymbol{x} = \boldsymbol{n}$. In the multi-class case, $\boldsymbol{x}$ is assigned membership in the class whose decision function $f_u(\boldsymbol{x})$ against its reference is maximum. Note that in acknowledging the presence of a reference class no assumptions are made about that class. While it is intuitive to think of the reference class as Gaussian noise, say, the reference class could be of any arbitrary distribution.

Below is an notional illustration of the CS-SVM concept for two dimensional data. Optimal separating hyperplanes for each class versus the noise-only reference class are found. Since the optimal hyperplane separating any two classes is unique [5], the optimal hyperplane for class i vs $\boldsymbol{n}$ will be different from the optimal hyperplane for class j vs $\boldsymbol{n}$. However, both hyperplanes are optimized against a common reference class. The decision function $f(\boldsymbol{x})$ for either signal-present class should reject the noise only case. Further, it is argued that $f_i(\boldsymbol{x})$ will be greater than $f_j(\boldsymbol{x})$ whenever $\boldsymbol{x}$ is associated with class i since $f_i(\boldsymbol{x})$ is optimal for class i and $f_j(\boldsymbol{x})$ is not.

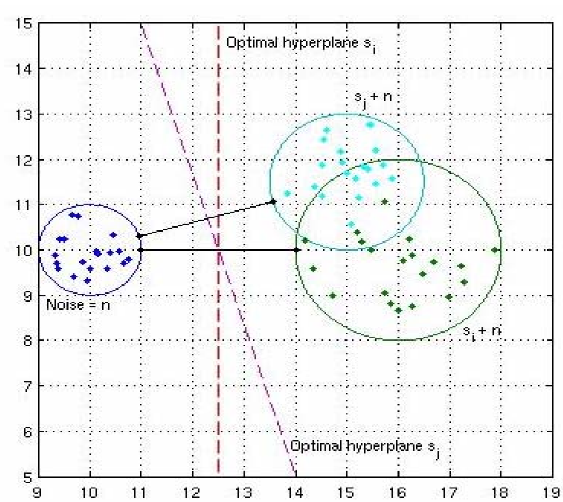


Figure 2: A geometric view of the optimal separating hyperplanes for two SVMs for class i and class j, respectively, in a 2-D decision space.

III. Experimental Results – Synthetic Data

To investigate the CS-SVM concept, several example cases using synthetic data were run. Figure 3 shows the training data and test data for two of the 2-D example cases tested. For these cases, a Gaussian radial basis function kernel was used such that $f(\boldsymbol{x}) = \sum_{i \in S} \alpha_i y_i \exp(-\|x - x_i\|^2 / 2\sigma^2) + b$ where S is the set of support vectors for which $\alpha_i > 0$. Training sets were produced separately for each signal-present class using Gaussian noise as the reference class such that

CS-SMV1: $T1 = \{(\boldsymbol{x}_1, y)\} = \{(\boldsymbol{s}_1 + \boldsymbol{n}, 1), (\boldsymbol{n}, -1)\}$ and
CS-SVM2: $T2 = \{(\boldsymbol{x}_2, y)\} = \{(\boldsymbol{s}_2 + \boldsymbol{n}, 1), (\boldsymbol{n}, -1)\}$

where $\boldsymbol{x}$, $\boldsymbol{s}$, and $\boldsymbol{n}$ were all 2-D vectors. Fifty samples of each case were generated in both training sets. Additionally, a training set suitable for a traditional binary SVM (B-SVM) was also generated again with fifty positive samples and fifty negative samples.

B-SVM: $T3 = \{(\boldsymbol{x}_3, y)\} = \{(\boldsymbol{s}_1 + \boldsymbol{n}, 1), (\boldsymbol{s}_2 + \boldsymbol{n}, -1)\}$

Each decision function $f_j(\boldsymbol{x})$ was then evaluated for test data consisting of 10000 samples from Class 1, 10000 samples from Class 2 and 10000 noise-only samples. The performance of the SVMs were evaluated using the following metrics. The results for the example cases are listed in Table 1.

$$P_{cc}(j) = \frac{\text{\# test samples from class } j \text{ where } f_j(x) > f_l(x)}{\text{Total \# of test samples from class } j} \quad \text{for all } j \neq l$$

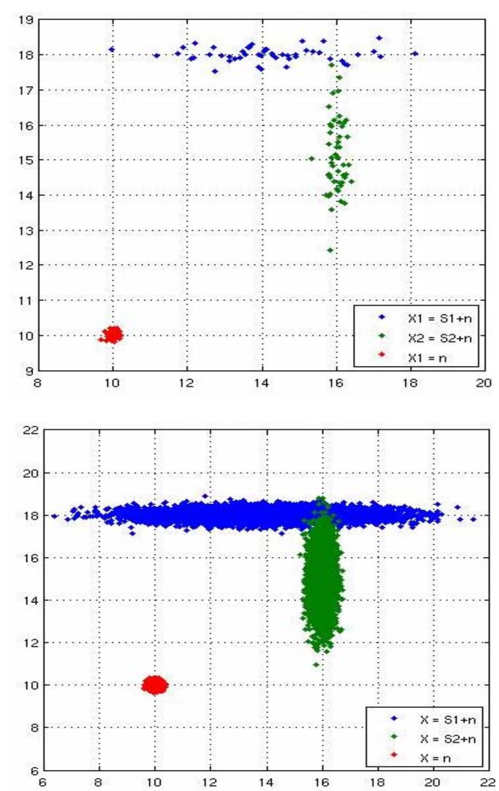
$$P_{miss}(j) = \frac{\text{\# test samples from class } j \text{ where } f_l(x) > f_j(x)}{\text{Total \# of test samples from class } j} \quad \text{where } l \neq j$$

$$P_{nse}(j) = \frac{\text{\# of noise-only test samples incorrectly classified as class } j}{\text{Total \# of noise-only test samples}}$$

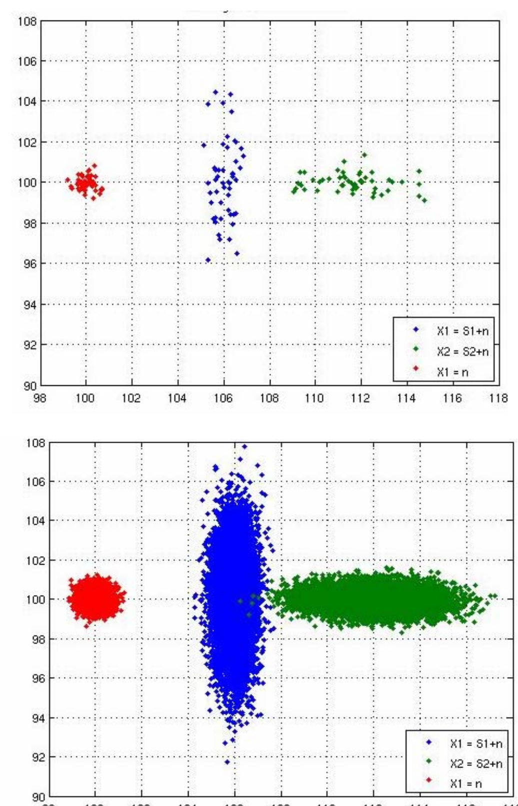
Overall, the CS-SVMs performed well. In Case 3 where the classes were (nearly) separable, the classification performance of the CS-SVM and B-SVM for the signal-present test data were comparable. However, the B-SVM, having knowledge of the noise-only condition, misclassified all of the noise-only test data as either class 1 or class 2. In Case 1 where the classes were significantly overlapped, the performance of the CS-SVM was again very good but support vector optimization (3) for the B-SVM failed. The resulting *f(x)* had no ability to separate the classes at all. Several values of the soft margin parameter *C* were tried without success.

Next, to explore the performance of the CS-SVM concept in a true multi-class setting, a synthetic 6 class case was considered. Training data and test data for the six class case are shown in figure 4. One SVM was constructed for each signal class versus noise-only using the training sets

CS-SVM u: $T_u = \{(\boldsymbol{x}_u, y)\} = \{(\boldsymbol{s}_u + \boldsymbol{n}, 1), (\boldsymbol{n}, -1)\}$ for $1 \leq u \leq 6$.



(a)



(b)

Figure 3: Training data (above) and test data (below) for 2 overlapped signal classes and noise-only reference. a) Case 1 and b) Case 3

Fifty positive and fifty negative examples were generated per class. A Gaussian radial basis function kernel was again used as *G* in (2) and (3). The performance of the CS-SVM for the 6 signal case is listed in Table 2.

TABLE 1: Performance of CS-SVM and binary SVM classifiers

Test Case	Classifier	P_{cc}	P_{miss}	P_{nse}
Case 1 *Overlapped*	CS-SVM1	0.9957	0.0043	0.0000
	CS-SVM2	0.9958	0.0042	0.0000
	B-SVM	*SV*	*Opt.*	*Failed*
Case 3 *Separated*	CS-SVM1	1.0000	0.0000	0.0000
	CS-SVM2	0.9938	0.0062	0.0000
	B-SVM	0.9988	0.0012	1.0 (N/A)

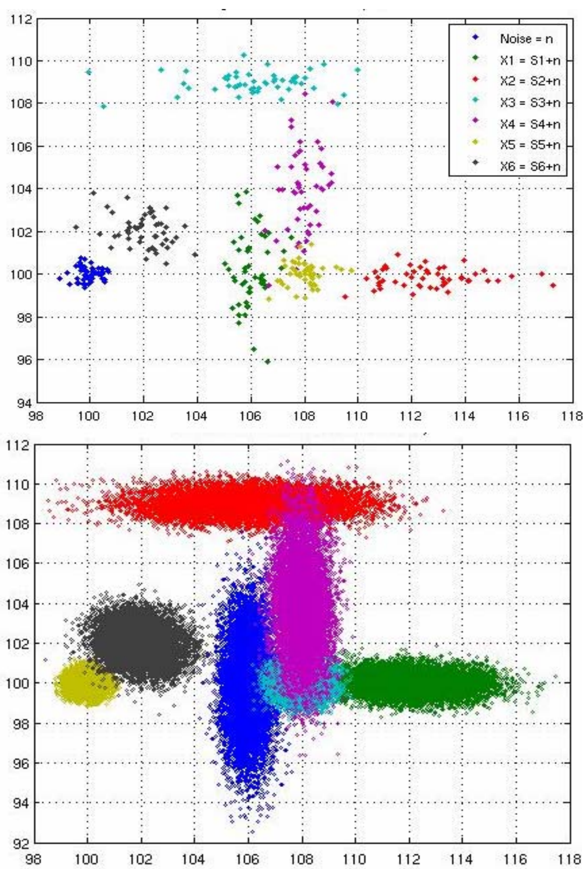


Figure 4: Training data (above) and test data (below) for six overlapped 2-D signal plus noise classes and the noise-only reference.

IV. CLASSIFICATION OF ODONTOCETE CLICKS

In the past several years there has been much interest and progress in acoustic monitoring, localization and tracking of marine mammals [14,15]. Acoustic monitoring has a number of benefits over visual monitoring. Chief among them are

TABLE 2: CS-SVM performance for the 6-class example case

Class	P_{cc}	P_{miss}	P_{nse}
1	0.9316	0.0684	0.0000
2	0.9268	0.0732	0.0000
3	0.9530	0.0470	0.0000
4	0.9142	0.0858	0.0000
5	0.6829	0.3171	0.0000
6	0.9878	0.0069	0.0053
Noise	--	0.0017	0.9983

increased area of coverage and the ability to operate over wider weather conditions and at night. A major drawback of acoustic monitoring is associating species information with the received vocalizations. However, recent field tests combining visual verification and digital recording tags with acoustic monitoring and localization have resulted in sets of “labeled” acoustic data [3]. These data are suitable for developing, training and testing classification algorithms.

Many toothed whale and dolphin species produce broadband click vocalizations. For species like pilot whales or dolphins, these clicks are just part of the animals' vocal repertoires which also include tonal whistles and sweeps. However, for other species like sperm whales and beaked whales, clicks are the primary sound they make. Given their involvement in multiple stranding events linked to mid-frequency sonar, the automated acoustic identification of beaked whales is of particular interest. Luckily for algorithm designers, beaked whale clicks appear to be quite distinctive.

Figure 5 shows the overlay of several clicks from Blainville's beaked whales (*Mesoplodon densirostris*) recorded during a September 2004 marine mammal tracking test at AUTEC [14]. As noted in [2], the clicks are actually FM sweeps. The level of similarity among the extracted clicks is striking. It should be noted that while these clicks all have similar peak amplitudes, they are not adjacent in time. They were selected across a 15 minute data segment. In fact, as beaked whales are often observed in groups of 3 or 4, there may even be calls from more than one animal present.

The first step in the design of a classification algorithm is to select a set of distinguishing features to represent the data such that the input vector to the classifier is $\boldsymbol{x} = [f_1 \;\; f_2 \; ... \; f_n]^T$. While the feature set should include as much information as possible, it should also be of reasonably low dimension because the amount of training data required grows with the dimension of the data. For *mesoplodon* clicks, the times between consecutive zero crossings were selected as the features. These features were chosen because a zero crossing detector is easy to implement and the periods between crossings capture the FM structure of the signal. Additionally, as is evident in figure 6, the measured periods of the first several zero crossings tend to cluster fairly tightly. In contrast,

the times between consecutive zero crossings for ambient noise data do not tend to cluster.

SVMs that discriminate between *mesoplodon* clicks set in ambient noise and ambient noise alone were developed using the periods of the first two, three, and four zero crossings as features. Each SVM was trained using 116 Blainsville's beaked whale clicks and 116 samples of ambient noise (fig. 7). The classifiers were then tested using 785 *mesoplodon* clicks taken from 2 different sites, located more than 15 Nmi apart, and 800 samples of ambient noise only take from one site. Note that the test data also included the training data. The classification performance versus ambient noise was excellent. Using the periods of the first four zero crossings as features, P_{cc}=0.985 and P_{nse}=0.010.

Next, SVMs were created for two other click-like signals that are commonly observed at AUTEC when *mesoplodon* clicks are present. Figure 8 shows ten clicks presumed to be from a pan-tropical spotted dolphin (*Stenella attenuata*) and a portion of ten man-made tracking pings used by the AUTEC range. The times between the first several consecutive zero crossings were again used as features with ambient noise used as the reference class. The SVMs for the *stenella* click were trained using 110 clicks and 110 ambient noise samples, and the SVMs for the tracking ping were trained using 120 pings and 120 ambient noise samples.

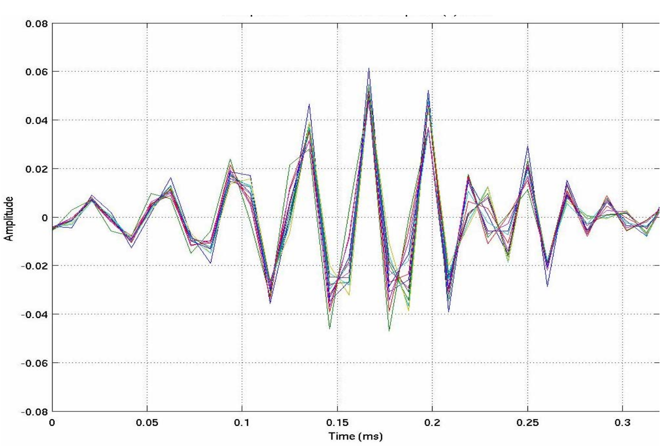


Figure 5: Twelve overlaid clicks from *Mesoplodon densirostris.*

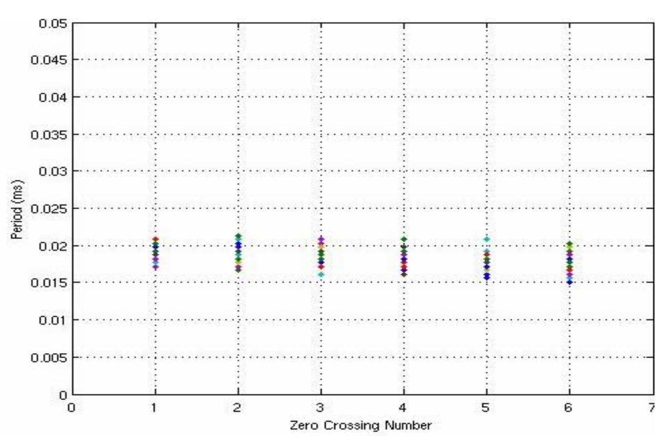


Figure 6: Times between consecutive zero crossing for 100 *mesoplodon* clicks.

Classification performance for each signal class individually against noise alone was again very good. The SVMs for *stenella* were tested using 1200 clicks and 1200 ambient noise samples. When the first 2 crossings were used P_{cc}=0.934 and P_{nse}=0.052, and when the first 3 crossings were used P_{cc}=0.876 and P_{nse}=0.042. The SVMs for the tracking pings were tested using 2000 pings with various amplitudes and Doppler shifts, and 2000 ambient noise samples. A P_{cc}=0.990 and a P_{nse}=0.070 were achieved using the periods of the first six zero crossings as the features.

The best SVM for each of the 3 classes individually were then combined to form a multi-class CS-SVM. Test input vectors $\boldsymbol{x}$ were assigned membership to class j^* according to $j^* = arg\ max(f_j(\boldsymbol{x}))$ or to the noise-only class if $max(f_j(\boldsymbol{x})) < 0$. The multi-class CS-SVM was tested using all the test data from each of the classes. The results are listed in Table 3. The greatest confusion among the classes occurred between the *stenella* click class and the tracking ping class. This is probably because the *stenella* class and the ping class are fairly close to each other in the chosen feature space (fig. 9). CS-SVM performance for *mesoplodon* clicks was excellent.

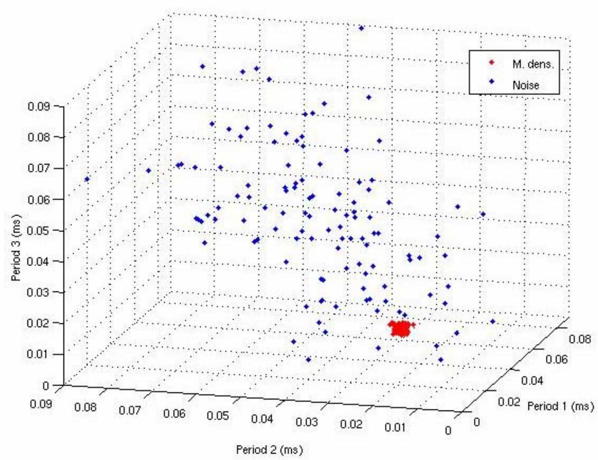


Figure 7: Scatter plot showing the distribution of the times between the first 3 zero crossing for 116 *mesoplodon* clicks and 116 ambient noise samples.

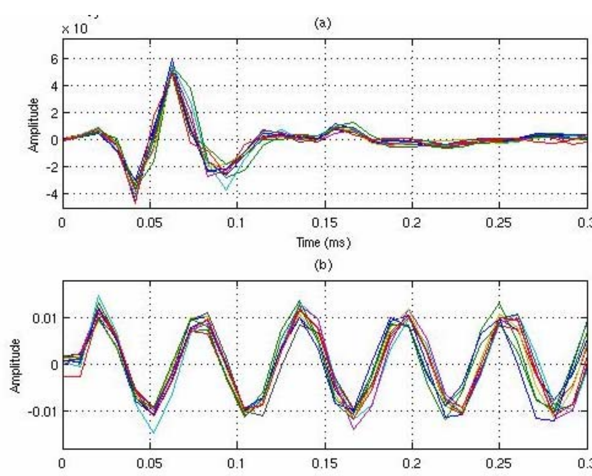


Figure 8: (a) Ten overlaid clicks believed to be from *Stenella attenuata*, and (b) the beginning portion of ten overlaid tracking pings.

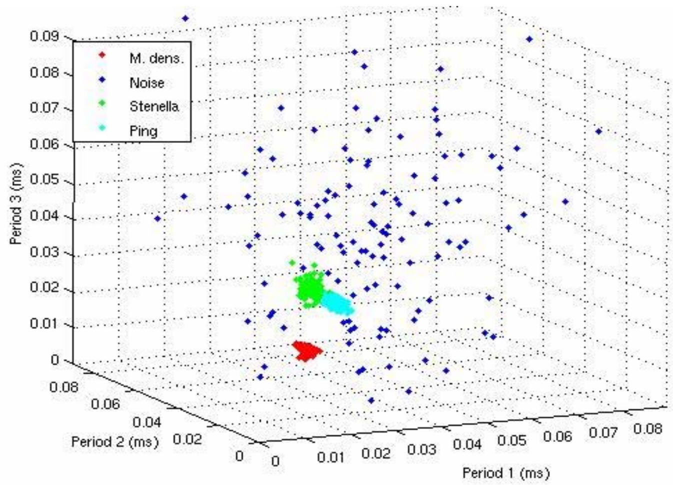


Figure 9: Scatter plot showing the distribution of the times between the first 3 zero crossing for 116 *mesoplodon* clicks , 116 ambient noise samples, 110 *stenella* clicks and 120 tracking pings. The *stenella* clicks and pings are fairly close together in this feature space.

TABLE 3: Performance of the CS-SVM for the 3 types of click waveforms

Test Data Set	P_{cc}	P_{miss}	P_{nse}
Mesoplodon (first 4 crossings)	0.9847	0.0013	0.0075
Stenella (first 3 crossings)	0.0.8817	0.0125	0.0408
Tracking Ping (first 6 crossings)	0.9495	0.0455	0.0250
Noise-only (all 3 sets)	--	0.0770	0.9230

V. CONCLUSION

This paper has presented a novel multi-class support vector machine classifier, the class-specific SVM. The new classifier consist of k binary SVMs where each SVM discriminates between one of k classes of interest and a common reference class. Test inputs are assigned membership in either the class whose decision function is maximized or the reference class if all decision function are negative. The CS-SVM concept was first demonstrated using several 2-dimensional synthetic examples. Then, a CS-SVM was created to classify click vocalizations from Blainville's beaked whale (*Mesoplodon densirostris).* The resulting classifier was able to reliably differentiate between *mesoplodon* clicks, delphinid clicks (from *Stenella attenuata*) and man-made tracking pings. The performance of the CS-SVM was excellent with over 98% of the test *mesoplodon* clicks correctly classified.

.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the on-going support of our sponsors, Dr. Mardi Hastings and Dr. Robert Gisiner, at the Office of Naval Research. Additionally, we would like to thank and acknowledge Diane Claridge and her team of observers from the Bahamas Marine Mammal Survey, Dr. Mark Johnson, Dr. Peter Tyack and their DTAG team from Woods Hole Oceanographic Institution and Dr. Joe Mobley of the University of Hawaii and his team of aerial observers for their collaborative efforts in collecting the visually verified recordings. Finally we would like to thank AUTEC for allowing us access to the range's considerable infrastructure and for continuing technical support. A special thanks to the following AUTEC personnel: Jose Arteiro, Marc Ciminello and Tom Szlyk.

REFERENCES

1 Johnson, M. and Tyack, P.L. (2003), “A digital acoustic recording tag for measuring the response of wild marine mammals to sound”, IEEE Journal of Ocean Engineering, vol. 28, pp 3-12.
2 Johnson, M., Madsen, P.T., et al., (2004) , “Beaked whales echolocate on prey”, Proceeding of the Royal Society London, B (Suppl.), Vol. 271, S383-386.
3 Moretti, et al, (2005) “Algorithms for passive acoustic detection and localization of marine mammals, including Blainville beaked whales (*Mesoplodon densirostris*), 2nd Int'l Conf. on Detection and Localization of Marine Mamals using Passive Acoustics, 16-18 November, Monoco.
4 Baggenstoss, P. (2002), www.npt.nuwc.navy.mil/csf/index.html
5 Vapnik, V. N. (1998), *Statistical Learning Theory*, New York, John Wiley & Sons, Inc.
6 Scholkopf, B., Smola, A. (2002), *Learning with Kernels*, Cambridge, MA, MIT Press.
7 Gunn, S.R., (1998), “Support Vector Machines for Classification and Regression”, Technical Report, School of Electronics, Univ. of Southampton.
8 www.mathworks.com/access/helpdesk/help/toolbox/optim/ug/quadprog.html.
9 Li, Z., Tang, S., and Yan, S., (2002), “Multi-Class SVM Classifier Based on Pairwise Coupling”, Proceedings of the First Int'l Workshop, SVM 2002, Niagara Falls, Canada, pp. 321.
10 Guemeur, Y., Elisseeff, A., Paugam-Moisey, H., (2000), “A new multi-class SVM based on a uniform convergence result”, Proceedings of the IEEE-INNS-ENNS International Joint Conf. on Neural Networks, Como, Italy, vol. 4, pp 183-188.
11 Weston, J. and Watson, C., (1998), “Multi-class Support Vector Machines”, Tech. Rep. CSD-TR-98-04, Royal Holloway, Univ. of London, Dept. of Computer Science.
12 Baggenstoss, P. M. (2003), “The PDF projection theorem and the class-specific method”, IEEE Transactions on Signal Processing, vol. 51(3), pp 685.
13 Bailey, A. (2001) *Class-dependent features and multicategory classification*, PhD Thesis, University of Southampton.
14 Jarvis, et al., (2003) “Passive monitoring and localization of marine mammals in open ocean environments using widely spaced hydrophones”, (Invited paper), 146th ASA Meeting, Austin TX.
15 Moretti, et al, (2005), “Passive acoustic detection and tracking of beaked whales in the Tongue of the Ocean”, Intergovernmental Conference on the Effects of Sound in the Ocean on Marine Mammals, 2-5 May, Lerici, Itally.